



faegredrinker.com

Jared Briant
Partner
Jared.briant@faegredrinker.com
+1 303 607 3588 direct

Faegre Drinker Biddle & Reath LLP
1144 15th Street, Suite 3400
Denver, Colorado 80202, USA
+1 303 607 3500 main
+1 303 607 3600 fax

July 24, 2024

VIA ECF

Honorable Sidney H. Stein
Daniel Patrick Moynihan
United States Courthouse
500 Pearl Street
New York, NY 10007

Re: *Authors Guild v. OpenAI, Inc.*, Case No. 23-cv-8292, and *Alter v. OpenAI, Inc.*, Case No. 23-cv-10211

Dear Judge Stein:

Microsoft submits this response to Plaintiffs' July 22, 2024 letter motion that seeks to compel documents that either do not exist or have already been produced. This is yet another example of Plaintiffs misrepresenting Microsoft's position and burdening the Court with discovery issues that do not need the Court's attention. Plaintiffs' letter motion focuses on a series of document requests regarding (1) Microsoft's training of ChatGPT, and (2) Microsoft's knowledge and awareness of OpenAI's activities in training ChatGPT. Microsoft has no documents regarding the first category because, as Microsoft has repeatedly informed Plaintiffs (and as Plaintiffs appear to admit), Microsoft did not train ChatGPT. And Plaintiffs are well aware that Microsoft has been producing responsive documents in its possession regarding knowledge and awareness of OpenAI's activities in training ChatGPT. Because there is no actual discovery dispute to resolve, Plaintiffs' letter motion should be denied.

Microsoft Has Produced Documents Regarding Its Knowledge and Awareness of OpenAI's Training Activities to the Extent They Exist.

Plaintiffs' assertion that Microsoft refuses to produce documents regarding OpenAI's training of ChatGPT is inaccurate and misleading. Plaintiffs should know from the parties' meet and confers and correspondence, the search terms Microsoft disclosed, and the thousands of documents Microsoft has produced to date, that Microsoft is producing documents regarding knowledge and awareness of OpenAI's activities in training ChatGPT in response to Plaintiffs' Request Nos. 17-18, 21, 24-25, 27, 40. In advance of producing these documents, counsel has (1) spoken with Microsoft personnel involved in the OpenAI relationship and/or potentially in possession of information received from and provided to OpenAI related to GPT models or ChatGPT; (2) searched and reviewed custodial documents for information related to the training of OpenAI's models, among many other topics; and (3) identified and reviewed documents from noncustodial locations related to Microsoft's relationship with OpenAI and/or information received from and provided to OpenAI related to GPT models or ChatGPT. In sum, Microsoft has

Honorable Sidney H. Stein

- 2 -

July 24, 2024

been conducting a thorough investigation both before and since responding to Plaintiffs' Requests for Production.

Microsoft's proposed search terms, custodians, and hit reports, provided to Plaintiffs on June 14, 2024, clearly demonstrate that Microsoft has searched for the term "train*" (in proximity to variations of "OpenAI" or "GPT"), and other training-related terms, and has produced responsive documents from these searches. Examples of these terms are below:

- (OpenAI OR OAI OR "Open AI" OR GPT* OR ChatGPT*) NEAR(10) train*
- (OpenAI OR OAI OR "Open AI" OR GPT* OR ChatGPT*) NEAR(50) pre-train*
- train* NEAR(50) licens* NEAR(50) (OpenAI OR OAI OR GPT* OR ChatGPT*)
- (OAI OR OpenAI OR "Open AI" OR GPT* OR ChatGPT*) NEAR(50) (index* NEAR(25) (data* OR train*))
- Colang* OR books1 OR books2

If responsive custodial documents exist, these terms will cover them, and Microsoft will diligently search for and produce any additional responsive material found through this ongoing custodial document review.¹ There are no further steps to take, and Plaintiffs have not identified any. Any further suggestion that Microsoft has not attempted to locate or produce documents regarding knowledge or awareness of OpenAI's training would be disingenuous. If what Plaintiffs are seeking here is instead a wide-ranging production of every comment, article, blog post, or other document that any of the more than 200,000 employees at Microsoft may have read regarding OpenAI's training of ChatGPT, such a request would be overly burdensome and disproportionate to the issues at stake in this case.

Microsoft Has Not Located Any Documents Regarding Microsoft's Training of ChatGPT Because Microsoft Did Not Train ChatGPT.

Plaintiffs misstate Microsoft's responses to Plaintiffs' RFP Nos. 22–23, 28–29, and 39, which each seek documents regarding Microsoft's—and not OpenAI's—training of ChatGPT. Microsoft did not state it would not search for documents in response to these requests but rather stated that "is not aware of any documents responsive to [these Requests] that are in Microsoft's possession, custody, or control." Microsoft's responses to these requests accurately state that Microsoft is not aware of responsive documents because "**Microsoft did not train ChatGPT.**" This statement—which remains true today—explained why, even four months ago when Microsoft served its responses, Microsoft stated to Plaintiffs that it was not aware of any responsive documents to requests seeking documents regarding **Microsoft's** training of ChatGPT. To the extent Plaintiffs understood this explanation to mean Microsoft was making no effort to locate documents responsive to the Requests, Plaintiffs are mistaken.

Plaintiffs conveniently gloss over the actual wording of these Requests in the letter motion, but a review of the actual language of Request Nos. 22, 28–29, and 39 reveals that the Requests

¹ Plaintiffs provided a counterproposal containing a request for additional search terms and custodians just last week, on July 18. As such, Plaintiffs' request for court intervention is also premature in light of these ongoing negotiations.

Honorable Sidney H. Stein

- 3 -

July 24, 2024

improperly assume, explicitly, by their use of “You” and “Your,” that **Microsoft** (and not OpenAI) conducted the activities set out in the Requests. (See, e.g., Letter Motion Ex. 1, Request No. 22 (“... YOU accessed commercial works of fiction and/or nonfiction, ... to train CHATGPT”), Request Nos. 28–29 (“... YOUR use of potentially copyrighted material in training CHATGPT”), and Request No. 39 (“... YOU purportedly used copyrighted material without permission in training CHATGPT”).) Microsoft informed Plaintiffs in its June 4 letter (Letter Motion Exhibit 4) that these requests do not seek documents regarding OpenAI’s training; they plainly seek documents regarding Microsoft’s training, which did not occur. Again, as stated in Microsoft’s objections and responses, as well as in correspondence and meet and confers, Microsoft did not train ChatGPT, and Microsoft cannot search for documents in response to nonsensical requests.² Plaintiffs cannot rewrite these fundamentally flawed requests now, and the Court need not do so for them. See *Fung-Schwartz v. Cerner Corp.*, No. 17-cv-0233-VSB-BCM, 2020 WL 4927485, at *8 n.8 (S.D.N.Y. Aug. 21, 2020) (court is not required to revise a party’s discovery request). Microsoft’s responses that there are no documents regarding Microsoft’s training of ChatGPT because it did not in fact train ChatGPT are complete and accurate. Nevertheless, Microsoft has continued to search for, and produce, documents related to OpenAI’s training activities.

Microsoft Has Not Located Any Documents Regarding the Use of Fiction or Nonfiction Works to Train ChatGPT.

Microsoft’s responses to Plaintiffs’ requests regarding the use of fiction works, nonfiction works, and copyrighted works to train ChatGPT are also complete and appropriate because Microsoft has not located any documents that are responsive to these requests. The Requests Plaintiffs identify do not seek documents regarding the content of the training data generally; they are instead directed specifically at “fiction works,” “nonfiction works,” “copyrighted works,” and the like. (See Letter Motion Ex. 1, at Request Nos. 22–23 (seeking documents regarding “commercial works of fiction [and/or] nonfiction, ... to train CHATGPT”), Request Nos. 28–29 (seeking documents regarding “use of potentially copyrighted material in training CHATGPT”), and Request Nos. 17–18, 21, 39 (requests directed specifically at fiction works, nonfiction works, and/or copyrighted material).) While Microsoft has produced documents regarding training generally and documents regarding knowledge or awareness of training, it has not located any documents regarding the use of “fiction works,” “nonfiction works,” or “copyrighted works” as training data.³ Yet again, Plaintiffs cannot now rewrite their request to seek documents they did not seek under Rule 34.

For these reasons, Plaintiffs’ letter motion seeking to compel production of documents that either do not exist or have already been produced should be denied.

² Despite serving these Requests that assume Microsoft trained ChatGPT, Plaintiffs now appear to tacitly concede that it was OpenAI that trained ChatGPT when Plaintiffs refer to “OpenAI’s training” in their letter motion.

³ Plaintiffs also rely on out of context language from a Microsoft/OpenAI agreement to argue that Microsoft *must* have documents regarding the substance of the data used by OpenAI to train ChatGPT, despite the fact that OpenAI has protected that information as a trade secret. As discussed, Microsoft searched for this information and it does not exist, and Plaintiffs have not identified any documents produced to date that suggest otherwise.

Honorable Sidney H. Stein

- 4 -

July 24, 2024

Respectfully submitted,

/s/ Jared B. Briant

Jared B. Briant

Counsel for Defendant Microsoft Corporation